DOI:
10.1007/978-3-319-25040-3_76

Citation for published version (APA):
Armstrong, J., & Brigo, D. (2016). Stochastic PDE projection on manifolds: Assumed-density and Galerkin filters. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9389, pp. 713-722). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 9389). SpringerVerlag Berlin Heidelberg. https://doi.org/10.1007/978-3-319-25040-3_76

Stochastic PDE projection on manifolds: Assumed-Density and Galerkin Filters

John Armstrong[1] and Damiano Brigo[2]

[1] Dept. of Mathematics, King's College London
[2] Dept. of Mathematics, Imperial College London
180 Queen's Gate, SW7 2AZ,
damiano.brigo@imperial.ac.uk

Abstract. We review the manifold projection method for stochastic nonlinear filtering in a more general setting than in our previous paper in Geometric Science of Information 2013. We still use a Hilbert space structure on a space of probability densities to project the infinite dimensional stochastic partial differential equation for the optimal filter onto a finite dimensional exponential or mixture family, respectively, with two different metrics, the Hellinger distance and the L^2 direct metric. This reduces the problem to finite dimensional stochastic differential equations. In this paper we summarize a previous equivalence result between Assumed Density Filters (ADF) and Hellinger/Exponential projection filters, and introduce a new equivalence between Galerkin method based filters and Direct metric/Mixture projection filters. This result allows us to give a rigorous geometric interpretation to ADF and Galerkin filters. We also discuss the different finite-dimensional filters obtained when projecting the stochastic partial differential equation for either the normalized (Kushner-Stratonovich) or a specific unnormalized (Zakai) density of the optimal filter.

1 The filtering problem in continuous time

The state of a system X evolves over time according to some stochastic process driven by a noise W. We cannot observe the state directly but we make an imperfect measurement Y which is also perturbed stochastically by random noise V. In a diffusion setting this problem is formulated as

$$dX_t = f_t(X_t)\,dt + \sigma_t(X_t)\,dW_t, \quad X_0, \qquad dY_t = b_t(X_t)\,dt + dV_t, \quad Y_0 = 0\ . \tag{1}$$

In these equations the unobserved state process $\{X_t, t \geq 0\}$ takes values in $\mathbb{R}^n$, the observation $\{Y_t, t \geq 0\}$ takes values in $\mathbb{R}^d$ and the noise processes $\{W_t, t \geq 0\}$ and $\{V_t, t \geq 0\}$ are two Brownian motions. The nonlinear filtering problem consists in finding the conditional probability distribution π_t of the state X_t given the observations up to time t and the prior distribution π_0 for X_0. Let us assume that X_0, and the two Brownian motions are independent. Let us also assume that the covariance matrix for V_t is invertible. We can then assume without any further loss of generality that its covariance matrix is the identity.

We introduce a variable a_t defined by $a_t = \sigma_t \sigma_t^T$. With these preliminaries, and a number of rather more technical conditions for which we refer to [9], one can show that π_t satisfies the Kushner–Stratonovich equation. We further suppose that the measure π_t is determined by a probability density p_t. A formal calculation then gives the following Stratonovich calculus version of the optimal filter stochastic PDE (SPDE) for the evolution of p:

$$dp_t = \mathcal{L}_t^* \, p_t \, dt - \frac{1}{2} \, p_t \, [|b_t|^2 - E_{p_t}\{|b_t|^2\}] \, dt + \sum_{k=1}^{d} p_t \, [b_t^k - E_{p_t}\{b_t^k\}] \circ dY_t^k \ . \quad (2)$$

We use Stratonovich calculus because we need the formal chain rule to hold when identifying the projected evolution from the projected right hand side of the equation, as we hint below after Equation (7).

Here $\mathcal{L}^*$ is the formal adjoint of $\mathcal{L}$ – the so-called forward diffusion operator for X, where the backward diffusion operator is defined by:

$$\mathcal{L}_t = \sum_{i=1}^{n} f_t^i \, \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} a_t^{ij} \, \frac{\partial^2}{\partial x_i \partial x_j}, \quad \mathcal{L}_t^* \phi = - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} [f_t^i \phi] + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} [a_t^{ij} \phi].$$

If the coefficients f and b are linear, σ is deterministic (and does not depend on X), and if the prior distribution is normal, this equation can be solved analytically to give the so-called Kalman Filter, where p is a Gaussian density. This Kalman filter reduces the problem to a vector SDE for the mean and a matrix SDE for the covariance matrix of the normal distribution. However, in general the optimal filter is not finite dimensional. We should point out that, in the general case, the preferred SPDE for the optimal filter is a SPDE for an unnormalized version q of the optimal filter density p. The Zakai equation for a specific unnormalized density $q_t(x)$ of the optimal filter reads, in Stratonovich form

$$dq_t = \mathcal{L}_t^* \, q_t \, dt - \frac{1}{2} \, q_t \, |b_t|^2 \, dt + \sum_{k=1}^{d} q_t \, [b_t^k] \circ dY_t^k \ , \quad q_0 = p_0,$$

see for example Eq. 14.31 in [1]. This is a linear Stochastic PDE and as such it is more tractable than the KS Equation. The reason why we still resort to KS will be clarified when we introduce the projection filter below. A general advantgage of the Zakai version is the possibility to derive a robust non-stochastic PDE for the optimal filter, see for example [10].

2 The projection method: From PDEs to ODEs

As we summarized previously in [3], the projection method can be understood abstractly as a technique to approximate the solution of a differential equation on a Riemannian manifold M. Given a vector field $\mathcal{X}$ defined on M, we wish to find the trajectory of a particle p as it flows along $\mathcal{X}$. We attempt to approximate this trajectory by choosing a submanifold Σ of M and using the Riemannian metric on M to project $\mathcal{X}$ applied to the current approximation p' in Σ onto the tangent space of Σ at p'. This gives rise to a tangent vector $\mathcal{X}'$ on Σ that is closest to the original L^2 tangent vector $\mathcal{X}$ in p'. One hopes that the trajectories of $\mathcal{X}'$ will be a good approximation for the trajectories of $\mathcal{X}$.

The approach becomes interesting for the filtering problem when one considers an infinite dimensional Hilbert manifold M where the exact stochastic PDE solution for the optimal filter evolves (in Stratonovich calculus), and a finite dimensional Σ where the projected stochastic ODE for the approximate filter will evolve. This idea was first sketched by Hanzon in [11] and fully developed in [8], [9], [2]. This is not only interesting for filtering. Indeed many standard approaches to the numerical solution of PDE's can be re-interpreted geometrically this way. Thus we will attempt to numerically solve the filtering problem by mapping the space of probability distributions into a Hilbert manifold and then projecting onto a finite dimensional submanifold. In fact the Hilbert manifolds we use will simply be Hilbert spaces.

3 Choice of Hilbert space structure

There are two obvious ways of embedding the state of our system as belonging in a Hilbert space. One can consider $\sqrt{p}$ which lies inside $L^2(\mathbb{R})$ or one can assume that p is itself square integrable and so lies inside $L^2(\mathbb{R})$. These two approaches give two different metrics on the space of probability distributions. The former yields the Hellinger metric, the latter we will call the direct L^2 metric.

Since the first approach requires no further assumptions on the integrability of p than p being integrable to one, the Hellinger metric immediately seems more attractive from a theoretical standpoint. Moreover, its definition can be extended to probability measures via their densities and it is invariant with respect to the base measure used to express densities of the two measures.

The direct L^2 metric is only defined on square integrable distributions and is not invariant under reparameterizations. However, it has one distinct advantage over the Hellinger metric: it is defined in terms of p rather than $\sqrt{p}$. Since the metric is bilinear in p, using the L^2 metric gives more convenient formulae for particular manifolds like mixture distributions than does the Hellinger metric, as we shall observe explicitly later. In [3] we observed that the direct metric also offers numerical advantages for mixture manifolds.

We should finally point out that the space of probability distributions is not a submanifold of $L^2(\mathbb{R})$. Fortunately we can view the stochastic PDE we wish to solve as an equation on the whole of $L^2(\mathbb{R})$ and so avoid the thorny question of defining a manifold structure on the space of probability measures, which is solved in [9] by introducing an enveloping manifold for the exponential case. A discussion on whether (the above Zakai version of) the optimal filter equation can be seen as a functional equation in L^2 is in the monograph [1]. More generally, the study of the infinite dimensional geometry for spaces of probability distributions is a broad field that has received increased attention over the last two decades, and we refer for example to [14] and [12].

4 Exponential and mixture submanifolds

Earlier research in [8], [9], [5] and [6] illustrated in detail how the Hellinger distance and the metric it induces on a finite dimensional exponential family, namely the Fisher metric, are ideal tools when using the projection onto exponential families of densities. The above references illustrate this by applying the above framework to the infinite dimensional stochastic PDE describing the optimal solution of the nonlinear filtering problem. The use of exponential families allows the correction step in the filtering algorithm to become exact, so that only the prediction step is approximated. Furthermore, and independently from the filtering application, exponential families and the Fisher metric are known to interact well thanks to a number of properties we will explain shortly.

We give now a summary of why the Fisher metric/Hellinger distance works well with exponential families and a summary of the Fisher-metric based projection filter. Section 4.2 will deal with the direct metric and mixtures families.

4.1 Exponential families

We use the following equivalent notations for multiple partial differentiation :

$$\frac{\partial^k}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} = \partial^k_{i_1,\cdots,i_k} \ .$$

Let $\{c_1, \cdots, c_m\}$ be scalar functions $c_i \ : \ \mathbb{R}^n \ \rightarrow \ \mathbb{R}$, $i = 1, 2, \cdots, m$ such that $\{1, c_1, \cdots, c_m\}$ are *linearly independent*, and assume that the convex set

$$\Theta_0 := \{\theta \in \mathbb{R}^m \ : \ \psi(\theta) = \log \int \exp[\theta^T c(x)]\, dx < \infty\} \ ,$$

has *non–empty interior*. Then

$$p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)],$$

where $\Theta \subseteq \Theta_0$ is open, is called an exponential family of probability densities. We define $E_\theta[\varphi] := \int \varphi(x) p(x, \theta) dx$. An important role in exponential families is played by differentiation of ψ. In fact for an exponential family ψ is infinitely differentiable in Θ and

$$E_\theta\{c_i\} = \partial_i\psi(\theta) =: \eta^i(\theta), \quad E_\theta\{c_i c_j\} = \partial^2_{ij}\psi(\theta) + \partial_i\psi(\theta)\, \partial_j\psi(\theta) \ ,$$

and more generally

$$E_\theta\{c_{i_1}\cdots c_{i_k}\} = \exp[-\psi(\theta)]\, \frac{\partial^k \exp[\psi(\theta)]}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} \ .$$

The Fisher information matrix satisfies

$$g_{ij}(\theta) = \partial^2_{ij}\psi(\theta) = \partial_i \eta^j(\theta) \ .$$

The quantities

$$(\eta^1, \cdots, \eta^m) \in \mathcal{E} = \eta(\Theta) \subset \mathbb{R}^m$$

form a coordinate system for the given exponential family. The two coordinate systems θ (canonical parameters) and η (expectation parameters) are related by diffeomorphism, and according to the above results the Jacobian matrix of the transformation $\eta = \eta(\theta)$ is the Fisher information matrix. The canonical parameters and the expectation parameters are *biorthogonal* w.r.t. the Fisher information metric.

We can now look at the particular shape taken by the Fisher metric projection for exponential families. We obtain

$$\Pi_\theta[v] = \sum_{i=1}^{m} [\sum_{j=1}^{m} g^{ij}(\theta) \ (E_\theta[vc_j] - E_\theta[v]E_\theta[c_j])] \ (c_i(\cdot) - E_\theta[c_i])p(\cdot, \theta). \tag{3}$$

The Fisher metric projection amounts to take covariance expectations of the function to be projected with the family exponents. The Fisher metric works well with exponential families essentially because in case of exponential families the square root amounts simply to add a 1/2 factor into the exponent of the family of density, and then differentiation of exponential functions is easy and regular.

4.2 Mixture Families

Besides exponential families, there is another general framework that is powerful in modeling probability densities, and this is the mixture family. Mixture distributions are ubiquitous in statistics and may account for important stylized features such as skewness, multi-modality and fat tails.

We define a mixture family as follows. Suppose we are given $m + 1$ fixed squared integrable probability densities, say $\underline{q} = [q_1, q_2, \ldots, q_{m+1}]^T$. Suppose we define the following space of probability densities:

$$p(x, \theta) = \theta_1 q_1(x) + \theta_2 q_2(x) + \cdots + \theta_m q_m(x) + (1 - \theta_1 - \cdots - \theta_m) q_{m+1}(x), \tag{4}$$

$$\theta \in \Theta, \quad \Theta = \{\theta : \ \theta_i \geq 0 \text{ for all } i, \quad \theta_1 + \cdots + \theta_m < 1\}.$$

For convenience, define $\hat{\theta}(\theta) := [\theta_1, \theta_2, \ldots, \theta_m, 1 - \theta_1 - \theta_2 - \ldots - \theta_m]^T$. With this definition, $p(x, \theta) = \hat{\theta}(\theta)^T \ q(x)$. If we consider the direct L^2 distance, the metric $h(\theta)$ that is induced on $p(x, \theta)$ and the related projection become very simple. Indeed, one can immediately check from the definition of h that for the mixture family we have tangent vectors and metric

$$\frac{\partial p(\cdot, \theta)}{\partial \theta_i} = q_i - q_{m+1}, \quad h_{ij}(\theta) = \int (q_i(x) - q_m(x))(q_j(x) - q_m(x)) dx =: h_{ij}$$

i.e., the L^2 direct metric (and matrix) does not depend on the specific point θ of the manifold. The same holds for the tangent space as we just saw. The L^2 projection is thus particularly simple:

$$\Pi_\theta[v] = \sum_{i=1}^{m}[\sum_{j=1}^{m} h^{ij}\ \langle v, q_j - q_{m+1}\rangle]\ (q_i - q_{m+1})\ . \tag{5}$$

We conclude by observing that, from the above calculations, the manifold for the direct metric that simplifies our projection equations drastically is the mixture choice. We analyzed numerical studies of the projection filter for the quadratic sensor with this direct metric - mixture setup and for the Hellinger-exponential setting in [3] under the special case of a scalar system for X, Y and with "q_i normal with mean μ_i and variance σ_i^2" and with "$c_i(x) = x^i$".

More generally, the motivation for considering these particular submanifolds is that, even in low dimensions, they allow us to reproduce many of the qualitative phenomena seen in the filtering problem. In particular we can produce highly skewed distributions and multi modal distributions. Many other possible choices of submanifold are worth consideration and are being investigated. In this respect, it is worth mentioning that in general there is no strict algorithmic method to select the manifold for a specific filtering problem, and this turns out to be a case by case matter. For example, in the quadratic sensor case one may expect a bimodal conditional density for the optimal filter, so one knows one will probably need about five parameters (two means, two standard deviations and a mixing parameter). As a general recipe, one can try a specific projection filter with a small manifold based on qualitative or heuristic considerations, as in the above-mentioned quadratic sensor case. Once this is done, one measures the L^2 norm of the projection residuals, and if this is large one may increase the manifold dimension until a sufficiently small projection residual norm is attained.

5 The projected equation

We now derive the direct L^2 projection filter for a general manifold M. Let M be an m dimensional submanifold of L^2 parameterized by $\theta = (\theta_1, \theta_2, \dots, \theta_m)$. Define $v_i = \frac{\partial p}{\partial \theta^i}$ so that $\{v_1, v_2, \dots v_m\}$ gives a basis for the tangent space of M at a point θ.

The direct L^2 metric induces a Riemannian metric $h_{ij}(\theta)$ on M. By projecting both sides of the Stratonovich equation for the evolution of p_t given above, we can obtain a stochastic differential for the evolution of the parameter θ.

To simplify the result, we introduce the following notation:

$$\gamma_t^0(p) := \frac{1}{2}\,[|b_t|^2 - E_p\{|b_t|^2\}]\ p, \quad \gamma_t^k(p) := [b_t^k - E_p\{b_t^k\}]p\ , \tag{6}$$

for $k = 1, \cdots, d$. Using the chain rule

$$dp(\theta_t) = \sum_{i=1}^{m} \frac{\partial p(\theta_t)}{\partial \theta^i} \circ d\theta_t^i \tag{7}$$

one can then show [2] that the projected equation for θ is equivalent to the stochastic differential equation:

$$\mathrm{d}\theta^i = \sum_{j=1}^{m} h^{ij} \left\{ \langle p(\theta), \mathcal{L} v_j \rangle \mathrm{d}t - \langle \gamma^0(p(\theta)), v_j \rangle \mathrm{d}t + \sum_{k=1}^{d} \langle \gamma^k(p(\theta)), v_j \rangle \circ \mathrm{d}Y^k \right\}. \quad (8)$$

Here $\langle \cdot, \cdot \rangle$ denotes the direct L^2 inner product. If preferred, one could instead project the Kushner–Stratonovich equation using the Hellinger metric instead. This yields the following stochastic differential equation [9]:

$$\mathrm{d}\theta^i = \sum_{j=1}^{m} g^{ij} \left(\langle \frac{\mathcal{L}^* p(\theta)}{p(\theta)}, v_j \rangle \mathrm{d}t - \langle \frac{1}{2} |b|^2, v_j \rangle \mathrm{d}t + \sum_{k=1}^{d} \langle b^k, v_j \rangle \circ \mathrm{d}Y^k \right) \quad (9)$$

It is now possible to explain why we resorted to the Kushner-Stratonovich (KS) Equation rather than the unnormalized but linear Zakai equation in deriving the projection filter. Consider the nonlinear terms in the KS Equation (2), namely

$$\frac{1}{2} p_t \, E_{p_t}\{|b_t|^2\} \, dt, \quad \sum_{k=1}^{d} p_t \, [-E_{p_t}\{b_t^k\}] \circ dY_t^k.$$

Consider first the Hellinger projection filter (9). By inspection, we see that there is no impact of the nonlinear terms in the projected equation. Therefore, projecting the Zakai equation would result in the same Hellinger projection filter.

Proposition 1. *The Hellinger projection takes care of dimensionality reduction and adds normalization as a bonus, without further approximation. Hellinger projection of either KS or the Zakai Eq. leads to the same projection filter given by Equation (9).*

This equivalence between KS and Zakai projection, however, is broken when we project according to the L^2 direct metric, obtaining the projection filter (8). For this filter we do have an impact of the nonlinear terms. In fact, it is easy to adapt the derivation of the L^2 direct filter to the Zakai equation, which leads to the filter

$$\mathrm{d}\theta^i = \sum_{j=1}^{m} h^{ij} \left\{ \langle p(\theta), \mathcal{L} v_j \rangle \mathrm{d}t - \langle \frac{1}{2} \, |b_t|^2 \; p(\theta), v_j \rangle \mathrm{d}t + \sum_{k=1}^{d} \langle b_t^k \; p(\theta), v_j \rangle \circ \mathrm{d}Y^k \right\} \quad (10)$$

which is clearly different from (8).

Proposition 2. *For the L^2 direct metric projection filter, the dimensionality reduction approximation coming with the projection does not take care of normalization and we obtain two different projection filters depending on whether we project the normalized KS Equation or the unnormalized Zakai Equation, leading to Equations (8) and (10) respectively.*

Since we aim at studying mostly the pure dimensionality reduction approximation, we use KS rather than Zakai, meaning that for the L^2 direct metric projection filter we will consider Equation (8) rather than (10). A numerical comparison of the two projection filters for the cubic sensor is under investigation.

6 Equivalence with ADF and Galerkin filters

The projection filter with specific metrics and manifolds can be shown to be equivalent to earlier filtering algorithms. We summarize the equivalence results here, starting from

ADF = ProjectionFilter(Hellinger, Exponential) (full details in [9]).

By computing the c-moments of the optimal filter $\hat{\eta}_i(t) = E[c_i(X_t)|\mathcal{Y}_t] = \int c_i(x)p_t(x)dx$ with p the optimal filter (2), one can write an equation for the $d\hat{\eta}_i(t)$ vector driven by dY. This will not be a closed vector differential equation, in that the right hand side will depend on the whole filter p_t and not just on its moments $\hat{\eta}_t$. However, if we *replace* the optimal filter p_t in the right hand side of this equation for $d\hat{\eta}(t)$ with the exponential density in the family with exponent c characterized by the expectation parameters $\hat{\eta}$, then we can close the differential equation and obtain a finite dimensional filter. This will not be the optimal filter but just an approximation, as the replacement is based on an arbitrary assumption. This approximation is called exponential assumed density filter (E-ADF). The resulting equation is

$$\begin{aligned} d\eta_t^i &= E_{\eta_t}\{\mathcal{L}_t\, c_i\}\, dt - \tfrac{1}{2}\,[\, E_{\eta_t}\{|b_t|^2\, c_i\} - E_{\eta_t}\{|b_t|^2\}\, \eta_t^i\,]\, dt \\ &\quad + \sum_{k=1}^{d} [\, E_{\eta_t}\{b_t^k\, c_i\} - E_{\eta_t}\{b_t^k\}\, \eta_t^i\,] \circ dY_t^k\ , \qquad i = 1, \cdots, m. \end{aligned} \tag{11}$$

Recall from our earlier section on exponential families that η's are an alternative coordinate system to η in the exponential manifold, so that the above equation for η can be seen as evolving in the exponential manifold. In fact, we can say more. In [9] we proved the following

Theorem 1. *The E-ADF (11) and the projection filter (9) on the same exponential family coincide. Forcing an exponential density on the right-hand-side of the exponent-moments equation results in the same filter as projecting the optimal filter onto the exponential family in Hellinger distance.*

This result is important because it shows that a heuristic approximation like the E-ADF can be justified in rigorous geometric terms by resorting to the Hellinger distance.

We now move to our second equivalence result:

Galerkin Filter = ProjectionFilter(Direct, Mixture) ([2] for details).

Our second equivalence result is that the projection filter in direct metric for simple mixture families is equivalent to an approximated filter derived via a Galerkin method, as first noticed in the preprint [7].

The basic Galerkin approximation is obtained by approximating the exact solution of the filtering SPDE (8) with a linear combination of basis functions $\phi_i(x)$, namely

$$\tilde{p}_t(x) := \sum_{i=1}^{\ell} \alpha_i(t)\phi_i(x). \tag{12}$$

Ideally, the ϕ_i can be extended to indices $\ell+1, \ell+2, \ldots, +\infty$ so as form a basis of L^2. The method can be sketched intuitively as follows. We could write the optimal filtering equation (2) as

$$\langle -dp_t + \mathcal{L}_t^* \, p_t \, dt - \gamma_t^0(p_t)\, dt + \sum_{k=1}^{d} \gamma_t^k(p_t) \circ dY_t^k \ , \xi \rangle = 0$$

for all smooth L^2 test functions ξ such that the inner product exists.

We replace this equation with the equation where p_t is replaced by $\tilde{p}_t$ in (12) and ξ is given by $\phi_1, \ldots, \phi_\ell$. Using the linearity of the inner product in each argument and integration by parts we obtain easily a stochastic ODE for the combinators $\alpha(t)$. We call this equation the Galerkin filter for ϕ.

Consider now the projection filter with the manifold (4) and the direct metric. The projection filter Equation (8) specializes to an equation that can be shown, by inspection, to be identical to the equation for the $d\alpha(t)$ coming from the Galerkin method if one sets

$$\ell = m{+}1, \ \alpha_i = \theta_i \text{ and } \phi_i = q_i - q_{m+1} \text{ for } i = 1, \ldots, m, \text{ and } \alpha_{m+1} = 1, \ \phi_{m+1} = q_{m+1}.$$

The choice of the simple mixture is related to a choice of the L^2 basis in the Galerkin method. A typical choice could be based on Gaussian radial basis functions, see for example [13]. We have thus sketched the proof of the following

Theorem 2. *For simple mixture families (4), the direct-metric projection filter (8) coincides with a Galerkin method where the basis functions are the mixture components q.*

However, this equivalence holds only for the simple mixture family (4). More complex mixture families, such as the one we used to analyze the quadratic sensor in [3], will not allow for a Galerkin-based filter and only the L^2 projection filter can be defined there. Note also that even in the simple case (4) our L^2 Galerkin/projection filter will be different from the Galerkin projection filter seen for example in [4], because we use Stratonovich calculus to project the Kushner-Stratonovich equation in L^2 metric.

References

1. Ahmed N. U. (1998). Linear and Nonlinear Filtering for Scientists and Engineers. World Scientific, Singapore.
2. Armstrong, J, Brigo D. (2013). Stochastic filtering via L^2 projection on mixture manifolds with computer algorithms and numerical examples. Submitted, preprint available at arXiv.org
3. Armstrong, J., & Brigo, D. (2013). Stochastic filtering by projection: The example of the quadratic sensor. In F. Nielsen, & F. Barbaresco (Eds.), Geometric Science of Information: First International Conference, GSI 2013, Paris, France (Lecture Notes in Computer Science; Vol. 8085, p. 685-692). Heidelberg: Springer.
4. Beard, R. and Gunther, J. (1997). Galerkin Approximations of the Kushner Equation in Nonlinear Estimation. Working Paper, Brigham Young University.
5. Brigo, D. Diffusion Processes, Manifolds of Exponential Densities, and Nonlinear Filtering, In: Ole E. Barndorff-Nielsen and Eva B. Vedel Jensen, editor, Geometry in Present Day Science, World Scientific, 1999
6. Brigo, D, On SDEs with marginal laws evolving in finite-dimensional exponential families, STAT PROBABIL LETT, 2000, Vol: 49, Pages: 127 – 134
7. Brigo, D. (2012). The direct L2 geometric structure on a manifold of probability densities with applications to Filtering. Available at arXiv.org
8. Brigo, D, Hanzon, B, LeGland, F, A differential geometric approach to nonlinear filtering: The projection filter, IEEE T AUTOMAT CONTR, 1998, Vol: 43, Pages: 247 – 252
9. Brigo, D, Hanzon, B, Le Gland, F, Approximate nonlinear filtering by projection on exponential manifolds of densities, BERNOULLI, 1999, Vol: 5, Pages: 495 – 534
10. Clark J. M. C. (1978). The design of robust approximations to the stochastic differential equations of nonlinear filtering, in: J.K. Skwirzynski, ed., Communication Systems and Random Process Theory, NATO Advanced Study Institute Series (Sijthoff and Noordhoff, Alphen aan den Rijn).
11. Hanzon, B. A differential-geometric approach to approximate nonlinear filtering. In C.T.J. Dodson, Geometrization of Statistical Theory, pages 219 – 223, ULMD Publications, University of Lancaster, 1987.
12. Khesin, B., Lennels, J., Misiolek, G., and Preston, S.C. Geometry of Diffeomorphism Groups, Complete integrability and Geometric statistics. Geom. Funct. Anal. Vol. 23 (2013) 334–366
13. Kormann, K., and Larsson, E. (2014). A Galerkin Radial Basis Function Method for the Schroedinger Equation. SIAM J. Sci. Comput., Vol. 35, No. 6, pp. A2832A2855
14. Pistone, G. and C. Sempi. An infinite-dimensional geometric structure on the space of all probability measures equivalent to a given one. Ann. Statist., 23 (1995), 1543–1561.